

Predictive Modeling Basics and Beyond

September 22, 2008

Introductions

Kate Hall, ASA, MAAA.
 Vice President, Solucia Inc.



Agenda Topics

- Data and Data Sources.
- Predictive Modeling:
 - Models and their uses;
 - Generating relative risk scores for underwriting.
- Predictive Modeling &
 - Renewal Underwriting;
 - New Business Underwriting.



Data

Types of Data for Predictive Modeling

- Claims
 - Medical
 - Pharmaceutical
- Eligibility
 - Enrollment
 - Census/Prospective Information
- Benefits
- Self-Reported
- DM and other Programs



Data Use Pitfalls

- Trusting data without diagnostics;
- Not knowing if data holes are systematic or sporadic;
- Not realizing how really tricky population data really is;
- Asking for too little;
- Not understanding, at least a little, about how the claims were paid.



Claims Data

Data Elements for Actuarial Modeling

- Individual Identifiers
- Diagnosis
- Services performed
- Unit counts
- Site of Service
- Dates
- Provider information
- Money
- Non-medical information



Claims Data

Data Elements for Predictive Model building

- Individual Identifiers;
- Diagnosis;
- Kind of Claim
 - Inpatient DRG's, Admits, Length of Stay;
 - Outpatient Facility Procedures, RevCode, APCs;
 - Professional Procedures, Specialty, Rx NDCs;
- Dates



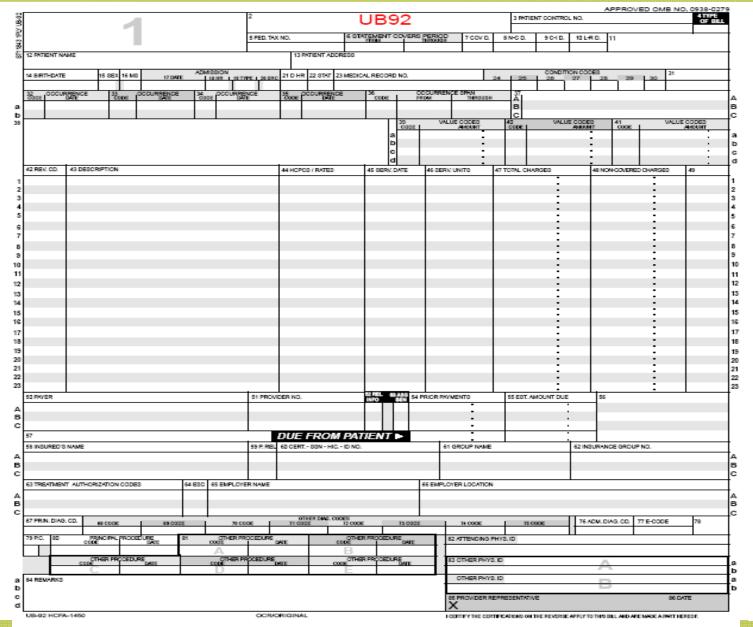
Claims information used for PM

Alphabet Soup

- UB 92
- CMS (HCFA) 1500
- Diagnosis
- DRG
- CPTs
- HCPCs



Sample Claim Form





Sample Claim Form

PLEASE DO NOT HCFA 1500 STAPLE IN THIS AREA HEALTH INSURANCE CLAIM FORM PICA 1a, INSURED'S LD, NUMBER (FOR PROGRAM IN ITEM 1) MEDICARE MEDICAID CHAMPUS CHAMPVA OTHER HEALTH BYAN (Modicare #) (Medicald #) (Spansor's SSM) (VA FIRE #1 I. INSURED'S NAME (Last Name, First Name, Middle Initial) 2. PATIENT'S NAME (Last Name, First Name, Middle Initial) 5. PATIENT'S ADDRESS (No., Street) 6. PATIENT RELATIONSHIP TO INSURED 7. INSURED'S ADDRESS (No., Street) Spouse B. PATIENT STATUS CITY STATE STATE Other TELEPHONE (INCLUDE AREA CODE) ZIP CODE ZIP CORE TELEPHONE (Include Area Code) Part-Time Student Student 10. IS PATIENT'S CONDITION FIELATED TO: 1. INSURED'S POLICY GROUP OR FECA NUMBER 9. OTHER INSURED'S NAME (Last Name, First Name, Middle Initial) a. EMPLOYMENT? (CURRENT OR PREVIOUS) a. INSURED'S DATE OF BIRTH a. OTHER INSURED'S POLICY OR GROUP NUMBER MM | DO YES b. OTHER INSURED'S DATE OF BIRTH b. AUTO ACCIDENT? PLACE (State) YES NO c. OTHER ACCIDENT? INSURANCE PLAN NAME OF PROGRAM NAME G. EMPLOYER'S NAME OR SCHOOL NAME YES NO 10d. RESERVED FOR LOCAL USE d. IS THERE ANOTHER HEALTH BENEFIT PLAN? d. INSURANCE PLAN NAME OR PROGRAM NAME READ BACK OF FORM BEFORE COMPLETING A SIGNING THIS FORM.

12. PATIENT'S OR AUTHORIZED PERSON'S SIGNATURE I authorize the release of any medical or other information necessary to process this claim. I also request payment of government benefits either to myself or to the party who accepts assignment below. YES NO # yes, return to and complete item 9 a-d.

13. INSURED'S OR AUTHORIZED PERSON'S SIGNATURE I authorize payment of medical benefits to the undersigned physician or supplier for services described below. ILLNESS (First symptom) OR INJURY (Accident) OR PREGNANCY(LMP) OF CURRENT: 16. DATES PATIENT UNABLE TO WORK IN CURRENT OCCUPATION
FROM MM | DD | YY
TO MM | DD | YY 18. HOSPITALIZATION DATES RELATED TO CURRENT SERVICES 17. NAME OF REFERRING PHYSICIAN OR OTHER SOURCE 17a. I.D. NUMBER OF REFERRING PHYSICIAN FROM " TO 19. RESERVED FOR LOCAL USE 20. OUTSIDE LAB? SCHARGES NO 22. MEDICAID RESUBMISSION CODE 21, DIAGNOSIS OR NATURE OF ILLNESS OR INJURY, (RELATE ITEMS 1,2,3 OR 4 TO ITEM 246 BY LINE) ORIGINAL REF. NO. 23. PRIOR AUTHORIZATION NUMBER ROCEDURES, SERVICES, OR SUPPLIE DATE(S) OF SERVICE Type of DIAGNOSIS RESERVED FOR (Explain Unusual Circumstances) CPT/HCPCS I MODIFIER S CHARGES Family Plan coe CODE 8 25. FEDERAL TAX I.D. NUMBER 28. PATIENT'S ACCOUNT NO. 27. ACCEPT ASSIGNMENT? ... IFor govt. claims, see back! 29. TOTAL CHARGE 29. AMOUNT PAID 30. BALANCE DUE YES 31. SIGNATURE OF PHYSICIAN OR SUPPLIER INCLUDING DEGREES OR CREDENTIALS NAME AND ADDRESS OF FACILITY WHERE SERVICES WERE RENDERED (Hother than home or office) 33. PHYSICIAN'S, SUPPLIER'S BILLING NAME, ADDRESS, ZIP CODE 8. PHONE # (I certify that the statements on the reverse apply to this bill and are made a part thereof.)



Individual Identifiers

- This is the first thing to get straight; everything else goes from here
- These should be unique but they are often so unique that they change with a new data system or when the beneficiary changes status or benefits or product
- They may have different lengths or formats in different systems
- You may need these to merge the information from the enrollment data to the information on the claims data looked for dropped records



Diagnosis Codes

- Diagnosis Codes state what is wrong with the patient or what might be wrong (rule out diagnosis and sometimes, admitting diagnosis).
- Diagnosis codes come, but seldom go. Keep track of changes annually. (One reason to use a grouper system: someone else responsible for the updates!).
- Diagnoses are structured mainly by disease neoplasm, infectious disease, maternity, disorders of the nervous system, but there are V codes as well which have other information.



Diagnosis formats (ICD-9)

- 3 digits are the most basic of the diagnoses
 - Diabetes Mellitus 250
- 4 digits are used for more description
 - 250.0 Diabetes mellitus without mention of complication
 - 250.1 Diabetes with ketoacidosis
 - 250.2 Diabetes with hyperosmolarity
 - 250.4 Diabetes with renal manifestations
 - 250.5 Diabetes with ophthalmic manifestations
 - 250.6 Diabetes with neurological manifestations
 - 250.7 Diabetes with peripheral circulatory disorders
 - 250.8 Diabetes with other specified manifestations
 - 250.9 Diabetes with unspecified complication



Diagnosis formats (ICD-9)

Sometimes a 5th digit is used for more detail

- Every 4 digit diabetes diagnosis can be further clarified into Type I or II, not stated as uncontrolled or uncontrolled, e.g.:
 - 250.60 Diabetes with neurological manifestations, type II or unspecified type, not stated as uncontrolled;
 - 250.61 Diabetes with neurological manifestations, type I [juvenile type], not stated as uncontrolled;
 - 250.62 Diabetes with neurological manifestations, type II or unspecified type, uncontrolled;
 - 250.63 Diabetes with neurological manifestations, type I [juvenile type], uncontrolled.



Diagnosis formats (ICD-9)

- Note some claims systems, say that if a 3-digit code has 4-digit codes that further describe it, then the 3-digit code is not acceptable for claim submission. If a 4-digit code has 5-digit codes that further describe it, then the 4-digit code is not acceptable for claim submission.
- For PM, all 5 is best but don't abandon your project because you only have 3.



Identification Algorithms (Groupers)

Examples:

- Commercial: ETGs, DXGs, GPIs, etc. Map diagnosis codes into groups that are predictive of resource utilization.
 May be better for one use than others.
- Self-developed: e.g. DMAA definitions/disease management industry definitions (see example, next page). Use-specific - may not be the best identification algorithm for underwriting applications.



Identification Algorithms (Groupers)

Medical Claims Only

ICD-9-CM CODES - ASTHMA	DESCRIPTION		
493.xx	Asthma		
493.01	Extrinsic Asthma, with mention of status asthmaticus or acute exacerbation or unspecified		
493.02	Asthma, with acute exacerbation		
493.10	Asthma, without mention of status asthmaticus or acute exacerbation or unspecified		
493.11	Asthma, with mention of status asthmaticus		
493.12	Asthma, with acute exacerbation		
493.9	Asthma, unspecified		
493.90	Asthma, unspecified without mention of status asthmaticus or acute exacerbation or unspecified		
493.91	Asthma, unspecified with mention of status asthmaticus		
493.92	Asthma, unspecified with acute exacerbation		
493.2	Chronic obstructive asthma		
493.20	Chronic obstructive asthma, without mention of status asthmaticus		
493.21	Chronic obstructive asthma, with mention of status asthmaticus		
493.22	Chronic obstructive asthma, with acute exacerbation		



Services Performed

- DRG
- CPT
- HCPC
- Revenue Code
- APC
- Other



Diagnosis Related Groups (DRG's)

- Inpatient payment method that combines diagnosis/diagnoses with services performed to give a case rate.
- Used for nearly all Medicare claims and some commercial claims.
- Need to watch for outlier claims.
- New MS-DRGs are effective and will be coming through in the data soon.



DRG Example

DRG	Type	Description	Weight
506	SURG	BURN W SKIN GRAFT OR INHAL INJ W CC OR SIG TRAUMA	4.7246
507	SURG	BURN W SKIN GRFT OR INHAL INJ W/O CC OR SIG TRAUMA	2.2603
508	MED	BURN W/O SKIN GRFT OR INHAL INJ W CC OR SIG TRAUMA	1.6171
509	MED	BURN W/O SKIN GRFT OR INH INJ W/O CC OR SIG TRAUMA	1.1338



CPT - Current Procedural Terminology

CPT's specifically describe services.

- For example 99201 is the code for a brief office visit with a new patient.
- The first digit of the code indicates the broad category -
 - 0 Anesthesia
 - 1-6 Surgery
 - 7 Radiology
 - 8 Pathology
 - 9 Medicine
- 99 Evaluation and Management (OV).



CPT

Sequential CPTs are usually very closely related.

These are all destruction of skin lesions:

17260 - lesion diameter 0.5 cm or less

17261 - lesion diameter 0.6 to 1.0 cm

17262 - lesion diameter 1.1 to 2.0 cm

17263 - lesion diameter 2.1 to 3.0 cm

17264 - lesion diameter 3.1 to 4.0 cm



CPT notes

- Modifiers are used to further explain the situation in which the procedure was performed.
- There are more than 40 modifiers.
- Some examples:
 - 26 Professional Component
 - 47 Anesthesia by Surgeon
 - 51 Multiple Procedures
 - 52 Reduced Service
 - 62 Two Surgeons
 - 63 Procedure Performed on Infants
 - 66 Surgical Team
 - 76 Repeat Procedure by Same Physician
 - 80 Assistant Surgeon



CPT tips

- CPT's are updated at least annually.
- Don't ignore modifiers when calculating average costs.
- Homegrown codes should be disappearing, but check to make sure.
- Check the totals spent by code to make sure there are no unexpected codes or there are no codes with a surprising amount of money, indicating a catch-all bucket.
- Keep track of terminated codes so they bucket correctly if they show up later.



HCPCS (Healthcare Common Procedure Coding System)

HCPCS includes all sorts of things:

- Ambulance, DME, Dental, Orthotics, Dental, Supplies, Screening Services, Injectable Drugs, New Technologies, etc, etc.
- They can be used for physician claims and facility claims.



HCPCS

Some codes are by definition temporary:

- C Codes are transitional pass throughs for outpatient services, stents, pacemakers, injectables.
- G Codes are for professional services advanced imaging, cancer screening, therapy, ESRD services.
- **S Codes** are for drugs, services, and supplies for which there are no national codes.



HCPCS

- When doing looking for services don't forget temporary HCPCS.
- Watch the unit measure, which can change./
- Link temporary services in prior year claims to final HCPC or CPT.
- Test data to see if codes are really what you think they are.
- Update crosswalk tables.



Revenue Codes

- Persistently used for Outpatient services although they are notoriously general.
 - For example the code 250, pharmacy, can be used for injections and aspirin.
 - Even if there are specific codes, such as 421 Therapy hourly charge, the bill may be for 420, therapy.



Revenue Codes

- If a claim has something more specific than a revenue code use it to categorize the claim.
- While revenue codes are mostly standardized, there are a lot of home grown categories that show up in the revenue code list.
- Worth testing any cross walk between CPT/HCPC and Revenue code to see what doesn't make sense or to find if it is a consistent mapping.



Enrollment Data

- Check that the individual identifier matches claims identifier.
- Know how the identifier changes as the member changes plans or groups or it will be hard to study individuals.
- Restate membership for retroactivity.
- Don't just use the membership as the denominator, merge membership information to claims to find out if there are anomalies or to get geography and demographics.



Non-medical "by" values

You may want to produce studies by something not inherently medical such as:

- Geography
- Group
- Product line
- Disease Management participation

DON'T use the claims information for this.

Use the enrollment files or the DM files.



Running risk scores

Data needed:

- Age
- Gender
- Diagnosis codes
- And possibly:
 - Procedure codes
 - NDC codes
 - Eligibility Category
 - Prior claims amounts
 - Self-reported data.



Running risk scores

- Model type
 - Concurrent
 - Prospective
 - Truncated
 - New types underwriting, high cost claimant, likelihood of hospitalization.
- For underwriting generally we want prospective and, potentially, truncated.



Risk Scores for Underwriting

- Some models predict what a member SHOULD cost.
- Underwriting cares what a member WILL cost.
- Possibly adjust commercial model to reflect member utilization patterns (high utilizer, etc.).



Running Risk Scores

- Input files:
 - Eligibility data
 - Claims or diagnosis data
 - Rx data
 - Model parameters (could be a file or could be selected via other methods).



Sample Input File

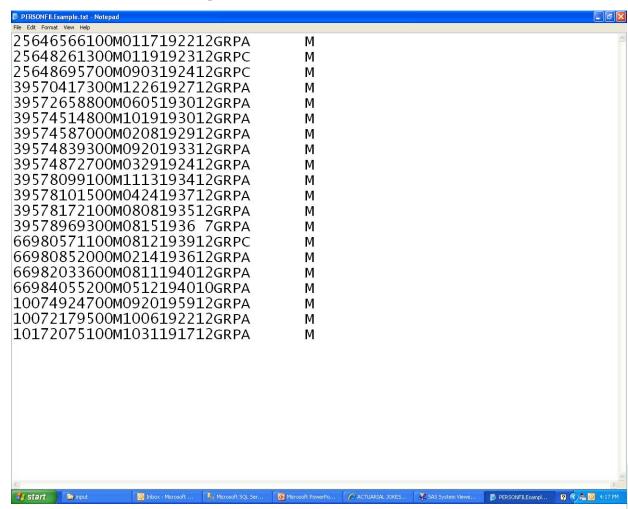
Diagnosis File Sample

```
diag_sample.TXT - Notepad
File Edit Format View Help
12577566100362.5
12577566100v04.8
12578261300998.1 881.00
12575261300998.1 881.00
12577261300998.1 881.00
12574261300998.1 881.00
12573261300427.3 428.0 401.9 4
12572261300427.3 428.0 401.9
12576261300709.9
12576261300998.1 881.00
12578261300709.9
12579261300709.9
12570261300709.9
12573261300216.3 238.2
12574261300428.0 427.31401.9 4
12575261300428.0 272.2 401.1 4
12576695700562.1 250.00530.8 1
12577695700562.1 250.00530.8 1
12578695700562.1 250.00530.8 1
12579695700562.1 250.00530.8 1
1257095700562.1 250.00530.8 1
12572695700562.1 250.00530.8 1
12572695700562.1 250.00530.8 1
12571695700562.1 250.00530.8 1
12571695700562.1 250.00530.8 1
12571695700562.1 250.00530.8 1
12571695700562.1 250.00530.8 1
10012695700410.7 486
                         112.4 1
# start in isksmart
              O Inbox - Microsoft ...
                               ☑ ( )♣ ◎ 4:12 PM
```



Sample Input File

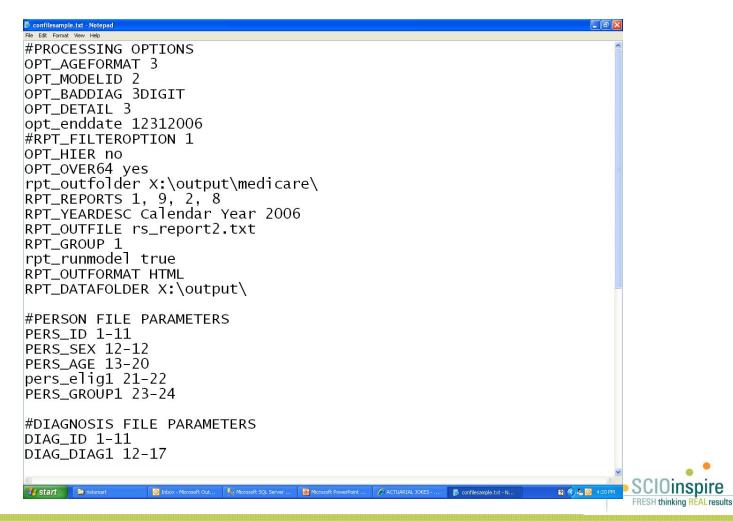
Member File Sample



FRESH thinking REAL results

Sample Input File

Parameter File Sample



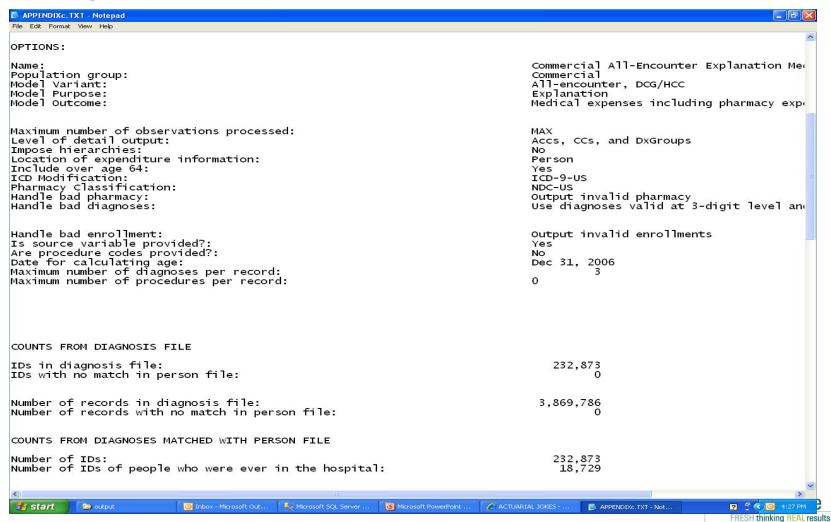
Running Risk Scores

Output Files

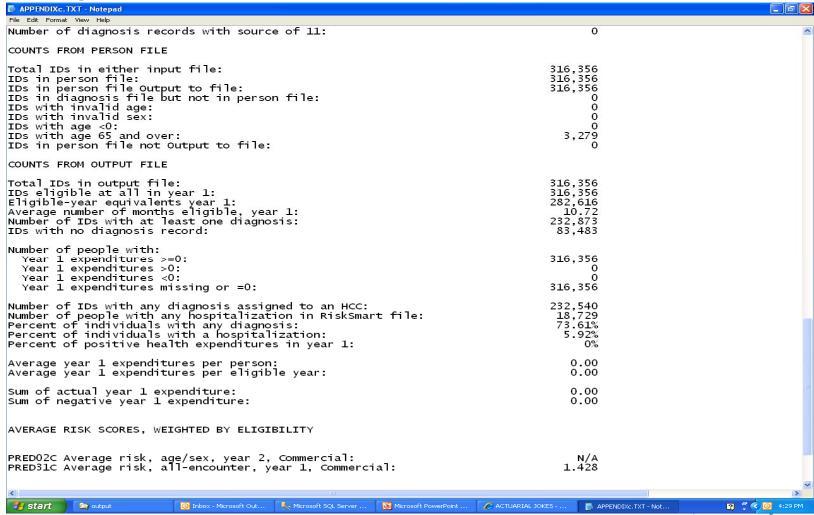
- Most important risk score file member level, diagnosis category flags, cost prediction, risk score.
- Statistics about the model run.
- Possibly reports.



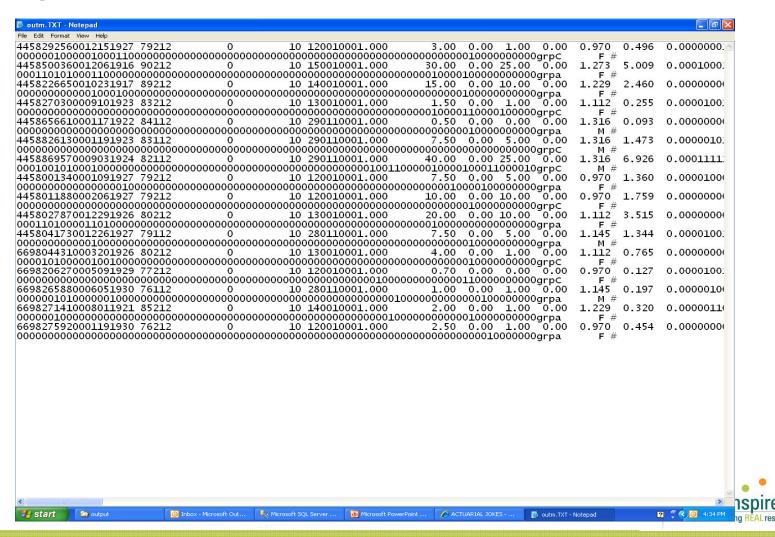
Output Statistics



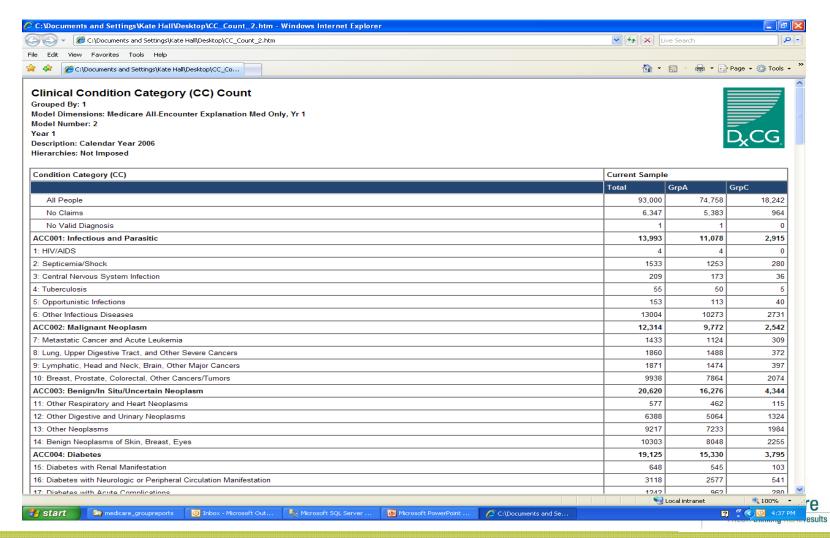
Output Statistics



Output Data



Output Reports



Questions?

Kate Hall, ASA MAAA

Solucia

220 Farmington Avenue

Suite 4

Farmington CT. 06032

860-676-8808 x 11

khall@soluciaconsulting.com

